

(19)日本国特許庁(JP)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開平11-66081

(43)公開日 平成11年(1999)3月9日

(51)Int.Cl.<sup>6</sup>

識別記号

F I

G 0 6 F 17/30

G 0 6 F 15/401

3 1 0 A

15/40

3 7 0 A

15/403

3 4 0 B

審査請求 未請求 請求項の数15 O L (全 12 頁)

(21)出願番号 特願平9-217619

(22)出願日 平成9年(1997)8月12日

(71)出願人 000102728

株式会社エヌ・ティ・ティ・データ

東京都江東区豊洲三丁目3番3号

(72)発明者 道坂 修

東京都江東区豊洲三丁目3番3号 エヌ・

ティ・ティ・データ通信株式会社内

(72)発明者 岩城 修

東京都江東区豊洲三丁目3番3号 エヌ・

ティ・ティ・データ通信株式会社内

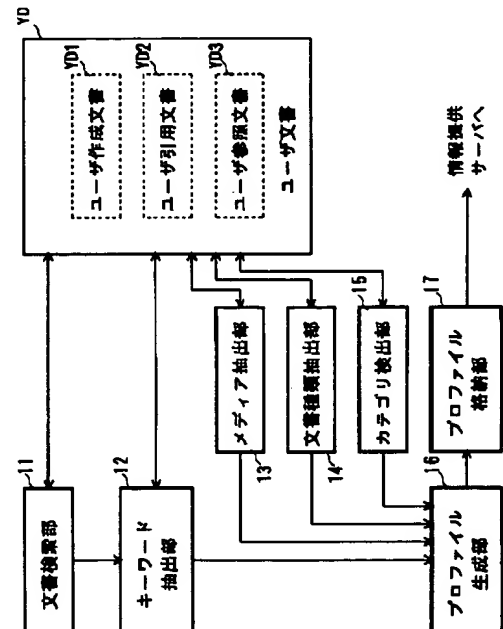
(74)代理人 弁理士 木村 満

(54)【発明の名称】 プロファイル取得システム、情報提供システム、プロファイル取得方法及び媒体

(57)【要約】

【課題】 情報提供システムに登録するためのユーザのプロファイルを取得し、ユーザの興味及び嗜好に適切に関連する情報の提供を可能とする。

【解決手段】 キーワード抽出部12は、ユーザ文書中の単語の出現頻度等に基づいてキーワードを抽出し、該キーワードに出現頻度、品詞及び文書構造に基づく重み付けを行い、プロファイル生成部16に与えてプロファイルに登録する。メディア抽出部13は、ユーザ文書のメディアの種類をプロファイルに登録する。文書種類抽出部14は、ユーザ文書の種類をプロファイルに登録する。カテゴリ検出部15は、ユーザ文書がカテゴリ化して管理されていれば、階層構造を示す情報をプロファイル生成部16に与え、キーワードをカテゴリ化させる。生成されたプロファイルは情報提供サーバのプロファイル格納部に登録され、このプロファイルに基づいて、情報が検索され、提供される。



## 【特許請求の範囲】

【請求項1】ユーザが作成した文書、ユーザが引用した文書及びユーザが参照した文書の少なくともいずれか1つを含むユーザ文書を検索する文書検索手段と、前記文書検索手段により検索されたユーザ文書から該ユーザ文書中の出現頻度に基づいてキーワードを抽出するキーワード抽出手段と、前記キーワード抽出手段により抽出されたキーワードに基づいて該ユーザのプロファイルを生成するプロファイル生成手段と、を具備することを特徴とするプロファイル取得システム。

【請求項2】前記キーワード抽出手段は、前記ユーザ文書における各キーワードの出現頻度、品詞、及び文書構造のうちの少なくともいずれかに基づき重み付けを前記キーワードに施す手段を含むことを特徴とする請求項1に記載のプロファイル取得システム。

【請求項3】前記プロファイル生成手段は、前記キーワード抽出手段により抽出されたキーワードについて、該キーワードの原始的意味を示すプリミティブを抽出し、該プリミティブを当該キーワードに属性として添付する手段を含むことを特徴とする請求項1又は2に記載のプロファイル取得システム。

【請求項4】前記文書検索手段により検索されたユーザ文書から、該ユーザ文書の主たるメディアの種類を特定するメディア抽出手段をさらに含み、前記プロファイル生成手段は、前記メディアの種類に基づいて該ユーザのプロファイルを生成する手段を含む、ことを特徴とする請求項1乃至3のうちのいずれか1項に記載のプロファイル取得システム。

【請求項5】前記文書検索手段により検索されたユーザ文書から、該ユーザ文書の主たる文書の種類を特定する文書種類抽出手段をさらに含み、前記プロファイル生成手段は、前記文書の種類に基づいて該ユーザのプロファイルを生成する手段を含むことを特徴とする請求項1乃至4のうちのいずれか1項に記載のプロファイル取得システム。

【請求項6】前記文書検索手段により検索されたユーザ文書から、該ユーザ文書のカテゴリライズに使用していたカテゴリ構造を検出するカテゴリ構造検出手段をさらに含み、且つ前記プロファイル生成手段は、前記カテゴリ構造に基づいて前記キーワードをカテゴリライズする手段を含むことを特徴とする請求項1乃至5のうちのいずれか1項に記載のプロファイル取得システム。

【請求項7】ユーザが作成した文書、ユーザが引用した文書及びユーザが参照した文書のうちの少なくともいずれかを含むユーザ文書に基づいてユーザに対応するプロファイルを取得するプロファイル取得手段と、前記プロファイル取得手段で取得されたプロファイルに基づいて、該プロファイルに対応する新たな情報を検索する情報検索手段と、

前記情報検索手段で新たな情報が検索された場合に、該新たな情報を提供する新規情報提供手段と、を具備することを特徴とする情報提供システム。

【請求項8】前記新規情報提供手段は、新たな情報を取得するための取得情報をユーザに通知する通知手段及び／又は新たな情報をユーザに転送する情報転送手段を含むことを特徴とする請求項7に記載の情報提供システム。

【請求項9】ユーザが作成した文書、ユーザが引用した文書及びユーザが参照した文書のうちの少なくともいずれかを含むユーザ文書に基づいてユーザに対応するプロファイルを取得するプロファイル取得手段と、前記ユーザから情報検索が要求された際に、前記プロファイル取得手段で取得された該ユーザに対応するプロファイルに基づいて、該プロファイルに対応する情報の重み付けを行う重み付け手段と、

前記ユーザから要求された情報を、前記重み付け手段による重み付けに従って検索し、該ユーザに提供する情報検索手段と、を具備することを特徴とする情報提供システム。

【請求項10】前記プロファイル取得手段は、前記ユーザ文書を検索する文書検索手段と、前記文書検索手段により検索されたユーザ文書から該ユーザ文書における出現頻度に基づいてキーワードを抽出するキーワード抽出手段と、前記キーワード抽出手段により抽出されたキーワードに基づいて該ユーザに対応するプロファイルを生成するプロファイル生成手段と、を含むことを特徴とする請求項7乃至9のうちのいずれか1項に記載の情報提供システム。

【請求項11】前記プロファイル取得手段は、所定のタイミングで、ユーザ文書に基づく前記プロファイルの取得更新を実行させる手段を含むことを特徴とする請求項7乃至10のうちのいずれか1項に記載の情報提供システム。

【請求項12】ユーザ文書を検索する文書検索ステップと、前記文書検索ステップにて検索されたユーザ文書からキーワードを抽出するキーワード抽出ステップと、前記キーワード抽出ステップにて抽出されたキーワードに基づいて該ユーザに対応するプロファイルを生成するプロファイル生成ステップと、を有することを特徴とするプロファイル取得方法。

【請求項13】ユーザが作成した文書、ユーザが引用した文書及びユーザが参照した文書のうちの少なくともいずれかを含むユーザ文書を検索する文書検索ステップと、前記文書検索ステップにて検索されたユーザ文書から出現頻度に基づいてキーワードを抽出するとともに、前記ユーザ文書における出現頻度、品詞、及び文書構造のう

10

20

30

40

50

ちの少なくともいずれかに基づく重み付けを前記キーワードに施すキーワード抽出ステップと、

前記文書検索ステップにて検索されたユーザ文書から、該ユーザ文書の主たるメディアの種類を特定するメディア抽出ステップと、

前記文書検索ステップにて検索されたユーザ文書から、該ユーザ文書の主たる文書の種類を特定する文書種類抽出ステップと、

前記文書検索ステップにより検索されたユーザ文書から、該ユーザ文書のカテゴリ化に使用していたカテゴリ構造を検出するカテゴリ構造検出ステップと、

前記キーワード抽出ステップにて抽出されたキーワードについて、該キーワードの原始的意味を示すブリミティブを抽出し、該ブリミティブを当該キーワードに属性として添付するステップ、前記メディアの種類に基づいて該ユーザに対応するプロフィールを生成するステップ、前記文書の種類に基づいて該ユーザに対応するプロフィールを生成するステップ、及び前記カテゴリ構造に基づいて前記キーワードをカテゴリ化するステップを含み、前記キーワード抽出ステップにて抽出されたキーワードに基づいて該ユーザに対応するプロフィールを生成するプロフィール生成ステップと、を有することを特徴とするプロフィール取得方法。

【請求項14】コンピュータを、ユーザが作成した文書、ユーザが引用した文書及びユーザが参照した文書のうちの少なくともいずれかを含むユーザ文書を検索する文書検索手段と、前記文書検索手段により検索されたユーザ文書から該ユーザ文書における出現頻度に基づいてキーワードを抽出するキーワード抽出手段と、

前記キーワード抽出手段により抽出されたキーワードに基づいて該ユーザに対応するプロフィールを生成するプロフィール生成手段と、して機能させるためのプログラムを記録した機械読み取り可能な媒体。

【請求項15】コンピュータを、ユーザが作成した文書、ユーザが引用した文書及びユーザが参照した文書のうちの少なくともいずれかを含むユーザ文書に基づいてユーザのプロフィールを取得するプロフィール取得手段と、

前記プロフィール取得手段で取得されたプロフィールに基づいて、該プロフィールに対応する新たな情報を検索する情報検索手段と、

前記情報検索手段で新たな情報が検索された場合に、該新たな情報を提供する新規情報提供手段と、して機能させるためのプログラムを記録した機械読み取り可能な媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、情報の共有及び情報の提供（情報共有／提供）を行うシステムにて個人

の興味及び嗜好等に応じた情報を選択的に提供するための技術に係り、特にユーザの個人情報であるプロフィールを自動的に取得するためのプロフィール取得システム、情報提供システム、プロフィール取得方法及び媒体に関する。

【0002】

【従来の技術】情報共有／提供を目的としたシステムでは、データベースに膨大な量の情報が管理されている。このため、システムの利用者すなわちユーザが、そのような膨大な量の情報の中から、所望する情報、つまりユーザの興味及び嗜好等に関連する情報を効率よく検索することは極めて困難である。そこで近年では、個人の興味及び嗜好等に対応する分野を事前に登録しておくことにより、個人の興味及び嗜好等に応じた情報を提供する情報提供サービスシステムが確立しつつある。

【0003】

【発明が解決しようとする課題】このような情報提供サービスシステムにおいて、望み通りの情報の提供を受けるためには、各ユーザは個人の興味及び嗜好等を示す情報（個人情報；プロフィール）を登録しなければならない。しかし、このような情報の選定及び登録は、それ自体が煩雑であり、また、必ずしも容易ではない。また、プロフィール（個人情報）の登録は、例えば情報提供サービスを行うサーバに接続するたびにその都度行う必要があり、さらに、煩わしいものとなっていた。

【0004】また、ユーザに提供される情報の品質は、登録されたプロフィールの内容及びその登録の仕方に依存する。従って、ユーザは適切にプロフィールを登録する必要がある。しかし、自分自身のプロフィールを適切に登録するためには、ある程度高度な知識及び経験を必要とし、一般の人には困難である。このため、この種の情報提供サービスでは、実質的に、ユーザ個人の興味及び嗜好等に沿った的確な情報を提供するには至っていない。

【0005】この発明は、上述した事情に鑑みてなされたもので、ユーザの適切なプロフィールを容易に取得するプロフィール取得システム及びプロフィール取得方法を提供することを目的とする。また、この発明は、ユーザの適切なプロフィールを取得し、該プロフィールの登録によりユーザの興味及び嗜好等に適切に関連する情報の提供を可能とするプロフィール取得システム、情報提供システム、プロフィール取得方法及び媒体を提供することを目的とする。

【0006】

【課題を解決するための手段】上記目的を達成するため、この発明の第1の観点によるプロフィール取得システムは、ユーザが作成した文書、ユーザが引用した文書及びユーザが参照した文書の少なくともいずれか1つを含むユーザ文書を検索する文書検索手段と、前記文書検索手段により検索されたユーザ文書から該ユーザ文書中

の出現頻度に基づいてキーワードを抽出するキーワード抽出手段と、前記キーワード抽出手段により抽出されたキーワードに基づいて該ユーザのプロファイルを生成するプロファイル生成手段と、を具備する。

【0007】この発明に係るプロファイル取得システムはユーザ文書を文書データベース等から検索し、該ユーザ文書中の出現頻度に基づいてキーワードを抽出して、該キーワードに基づいて該ユーザに対応するプロファイルを生成する。従って、ユーザがよく使用する或いはよく参照する用語やフレーズがキーワードとして自動的に取得・設定される。従って、ユーザは、カテゴリやキーワードを設定する必要がなく、簡単にプロファイルを取得・設定できる。

【0008】前記キーワード抽出手段は、前記ユーザ文書における出現頻度、品詞、及び文書構造のうちの少なくともいずれかに基づく重み付けを前記キーワードに施す手段を含んでいてもよい。このような構成とすることにより、単なる出現頻度に基づくものよりも、より適切なキーワードを抽出することができる。

【0009】前記プロファイル生成手段は、前記キーワード抽出手段により抽出されたキーワードについて、該キーワードの原始的意味を示すプリミティブを抽出し、該プリミティブを当該キーワードに属性として添付する手段を含んでいてもよい。

【0010】前記文書検索手段により検索されたユーザ文書から、該ユーザ文書の主たるメディアの種類を特定するメディア抽出手段をさらに含み、且つ前記プロファイル生成手段は、前記メディアの種類に基づいて該ユーザに対応するプロファイルを生成する手段を含んでいてもよい。

【0011】前記文書検索手段により検索されたユーザ文書から、該ユーザ文書の主たる文書の種類を特定する文書種類抽出手段をさらに含み、且つ、前記プロファイル生成手段は、前記文書の種類に基づいて該ユーザに対応するプロファイルを生成する手段を含んでいてもよい。

【0012】前記文書検索手段により検索されたユーザ文書から、該ユーザ文書のカテゴリライズに使用していたカテゴリ構造を検出するカテゴリ構造検出手段をさらに含み、且つ前記プロファイル生成手段は、前記カテゴリ構造に基づいて前記キーワードをカテゴリライズする手段を含んでいてもよい。

【0013】この発明の第2の観点による情報提供システムは、ユーザが作成した文書、ユーザが引用した文書及びユーザが参照した文書のうちの少なくともいずれかを含むユーザ文書に基づいてユーザに対応するプロファイルを取得するプロファイル取得手段と、前記プロファイル取得手段で取得されたプロファイルに基づいて、該プロファイルに対応する新たな情報を検索する情報検索手段と、前記情報検索手段で新たな情報が検索された場

合に、該新たな情報を前記ユーザに提供する新規情報提供手段と、を具備する。

【0014】この情報提供システムは、ユーザ文書に基づいてユーザに対応するプロファイルを取得し、該プロファイルに基づき、情報を検索して前記ユーザに提供する。従って、ユーザが個別にプロファイル等を入力しなくても、適切なプロファイルを自動的に取得及び登録して、ユーザ個人の興味及び嗜好等に適切に関連する情報を提供することができる。

10 【0015】前記新規情報提供手段は、ユーザに新たな情報を取得するための取得情報を通知する通知手段を含んでいてもよい。前記新規情報提供手段は、ユーザに新たな情報を転送する情報転送手段を含んでいてもよい。

【0016】この発明の第3の観点による情報提供システムは、ユーザが作成した文書、ユーザが引用した文書及びユーザが参照した文書のうちの少なくともいずれかを含むユーザ文書に基づいてユーザに対応するプロファイルを取得するプロファイル取得手段と、前記ユーザから情報検索が要求された際に、前記プロファイル取得手段で取得された該ユーザに対応するプロファイルに基づいて、該プロファイルに対応する情報の重み付けを行う重み付け手段と、前記ユーザから要求された情報を、前記重み付け手段による重み付けに従って検索し、該ユーザに提供する情報検索手段と、を具備する。

【0017】また、この情報提供システムは、ユーザ文書に基づいてユーザに対応するプロファイルを取得し、該プロファイルに基づく情報の重み付けを行って、情報を検索して前記ユーザに提供する。従って、ユーザが個別にプロファイル等を入力しなくても、適切なプロファイルを自動的に取得し、しかも、それに基づいた情報の重み付けを行って、ユーザ個人の興味及び嗜好等に適切に関連する情報を提供することができる。

【0018】前記プロファイル取得手段は、前記ユーザ文書を検索する文書検索手段と、前記文書検索手段により検索されたユーザ文書から該ユーザ文書における出現頻度に基づいてキーワードを抽出するキーワード抽出手段と、前記キーワード抽出手段により抽出されたキーワードに基づいて該ユーザに対応するプロファイルを生成するプロファイル生成手段と、を含んでいてもよい。

40 【0019】前記プロファイル取得手段は、所定のタイミングで、ユーザ文書に基づく前記プロファイルの取得更新を実行させる手段を含んでいてもよい。

【0020】この発明の第4の観点によるプロファイル取得方法は、ユーザ文書を検索する文書検索ステップと、前記文書検索ステップにて検索されたユーザ文書からキーワードを抽出するキーワード抽出ステップと、前記キーワード抽出ステップにて抽出されたキーワードに基づいて該ユーザに対応するプロファイルを生成するプロファイル生成ステップと、を有する。

50 【0021】この発明の第5の観点によるプロファイル

取得方法は、ユーザが作成した文書、ユーザが引用した文書及びユーザが参照した文書のうちの少なくともいずれかを含むユーザ文書を検索する文書検索ステップと、前記文書検索ステップにて検索されたユーザ文書から出現頻度に基づいてキーワードを抽出するとともに、前記ユーザ文書における出現頻度、品詞、及び文書構造のうちの少なくともいずれかに基づく重み付けを前記キーワードに施すキーワード抽出ステップと、前記文書検索ステップにて検索されたユーザ文書から、該ユーザ文書の主たるメディアの種類を特定するメディア抽出ステップと、前記文書検索ステップにて検索されたユーザ文書から、該ユーザ文書の主たる文書の種類を特定する文書種類抽出ステップと、前記文書検索ステップにより検索されたユーザ文書から、該ユーザ文書のカテゴリライズに使用していたカテゴリ構造を検出するカテゴリ構造抽出ステップと、前記キーワード抽出ステップにて抽出されたキーワードについて、該キーワードの原始的意味を示すブリミティブを抽出し、該ブリミティブを当該キーワードに属性として添付するステップ、前記メディアの種類に基づいて該ユーザに対応するプロフィールを生成するステップ、前記文書の種類に基づいて該ユーザに対応するプロフィールを生成するステップ、及び前記カテゴリ構造に基づいて前記キーワードをカテゴリライズするステップを含み、前記キーワード抽出ステップにて抽出されたキーワードに基づいて該ユーザに対応するプロフィールを生成するプロフィール生成ステップと、を有する。

【0022】この発明の第6の観点による媒体は、コンピュータを、ユーザが作成した文書、ユーザが引用した文書及びユーザが参照した文書のうちの少なくともいずれかを含むユーザ文書を検索する文書検索手段と、前記文書検索手段により検索されたユーザ文書から該ユーザ文書における出現頻度に基づいてキーワードを抽出するキーワード抽出手段と、前記キーワード抽出手段により抽出されたキーワードに基づいて該ユーザに対応するプロフィールを生成するプロフィール生成手段と、して機能させるためのプログラムを記録している。

【0023】この発明の第7の観点による媒体は、コンピュータを、ユーザが作成した文書、ユーザが引用した文書及びユーザが参照した文書のうちの少なくともいずれかを含むユーザ文書に基づいてユーザのプロフィールを取得するプロフィール取得手段と、前記プロフィール取得手段で取得されたプロフィールに基づいて、該プロフィールに対応する新たな情報を検索する情報検索手段と、前記情報検索手段で新たな情報が検索された場合に、該新たな情報を提供する新規情報提供手段と、して機能させるためのプログラムを記録している。

【0024】

【発明の実施の形態】以下、この発明の実施の形態を図面を参照して説明する。図1〜図3を参照して、この発明の第1の実施の形態に係るプロフィール取得システム

を説明する。

【0025】図1は、この発明の第1の実施の形態に係るプロフィール取得システムの構成を模式的に示している。

【0026】図1に示すプロフィール取得システムは、文書検索部11、キーワード抽出部12、メディア抽出部13、文書種類抽出部14、カテゴリ検出部15、プロフィール生成部16、及びプロフィール格納部17を備えている。

【0027】文書検索部11は、ユーザに関連し且つ当該コンピュータシステムの記憶装置内のユーザ文書YD又は当該コンピュータシステムから参照し得るユーザ文書YDを検索する。ユーザ文書YDには、ユーザ自身が過去に作成したユーザ作成文書YD1、ユーザが過去に引用したことがあるユーザ引用文書YD2及びユーザが過去に参照したことがあるユーザ参照文書YD3等が含まれる。

【0028】なお、この検索のため、例えば、このコンピュータ装置の文書作成機能（ワードプロセッサアプリケーション）は、作成した文書に文書情報として、作成者、作成日付、編集者、編集日付等の情報を含む文書情報を付す。また、文書作成及び参照・引用・読込等の履歴、参照・引用・読込を行った文書へのリンクのリスト等を保持する。文書検索部11は、例えば、コンピュータ内に保存されている文書を検索することにより、或いは、履歴を参照して、文書をアクセスすることにより、リンクを辿って検索した文書の文書情報を検索することにより、ユーザ文書YDを検索することができる。

【0029】キーワード抽出部12は、上述したユーザ作成文書YD1、ユーザ引用文書YD2及びユーザ参照文書YD3等のユーザ文書YDからキーワードを抽出し、プロフィール生成部16に与えてプロフィールに登録する。すなわち、キーワード抽出部12は、例えばTF（Term Frequency）及び／又はIDF（Inverse Document Frequency）法を用い、ユーザ文書YD内の単語の出現頻度に基づいてキーワードを抽出し、抽出したキーワードに該出現頻度に基づく重み付けを行う。例えば、ユーザ文書YD中に「特許」というワードが1000回出現し、「出願する」というワードが500回あらわれた場合には、「特許」の重みW1を大きくし、「出願する」の重みW2を相対的に小さくする。また、該キーワードには、品詞の種類に基づく重み、文書構造に基づく重みを加算する。例えば、「特許」は名詞なので重みW3を大きくし、「出願する」は動詞なので重みW4を小さくする。各キーワードの最終的な重みは、例えば、各キーワードについての各重みの積になる。例えば、「特許」と「出願する」の重みは、W1・W3と、W2・W4になる。さらに、該キーワードについて、シソーラス辞書を参照して、原始的意味を示すブリミティブを抽出し、該ブリミティブをキーワードに属性として付加す

る。例えば、「出願する」は、「願う」が属性として付加される。

【0030】メディア抽出部13は、ユーザ文書Y Dのメディア（新聞、専門新聞、一般雑誌、専門誌、等）の種類を調べ、特定のメディアが支配的であれば、それを抽出して、プロファイル生成部16に与え、プロファイルに登録する。文書種類抽出部14は、ユーザ文書Y Dの文書の種類（メール、雑誌記事、新聞記事）の分布を調べ、特定の文書種類が支配的であれば、それを抽出して、プロファイル生成部16に与え、プロファイルに登録する。

【0031】カテゴリ抽出部15は、ユーザによるユーザ文書Y Dの管理構造を調べ、フォルダ等の階層構造にカテゴリライズして管理されていれば、該階層構造を示す情報をプロファイル生成部16に与え、プロファイルにおけるキーワードを該階層構造を採用してカテゴリライズさせる。

【0032】プロファイル生成部16は、キーワード抽出部12、メディア抽出部13、文書種類抽出部14及びカテゴリ抽出部15から与えられる、キーワード、重み、属性及び階層構造の情報に基づいてユーザ個人のプロファイルを作成して、プロファイル格納部17に格納する。プロファイル格納部17は、プロファイル生成部16により生成されたユーザのプロファイルを格納する。このプロファイル格納部17に格納されたユーザプロファイルは、必要に応じて情報提供サーバ等に転送される。

【0033】次に、図2及び図3に示すフローチャートを参照して、図1のように構成されたプロファイル取得システムの動作を説明する。

【0034】ここでは、前提として、(a) ユーザ自身が作成した文書、及び(b) ユーザが興味に基づいて、引用及び参照するために収集した文書の実体あるいはその実体をたどって入手することを可能とするリンク（具体的には例えばインターネットの場合のURL（Universal Resource Locator）等）などがコンピュータ内に存在するものとする。

【0035】これらのユーザ自身が作成したユーザ作成文書Y D1、引用したユーザ引用文書Y D2及び参照したユーザ参照文書Y D3等のユーザ文書Y Dから、プロファイルを作成・生成して取得するシステムが、この発明の実施の形態によるプロファイル取得システムである。

【0036】プロファイルの取得は、図2に示す過程で行う。

(1) 文書検索部11が、ユーザの検索端末等のコンピュータから、ユーザが作成したユーザ作成文書Y D1、引用したユーザ引用文書Y D2及び参照したユーザ参照文書Y D3等を含むユーザ文書Y Dを検索する（ステップS1）。このとき、検索したユーザ文書Y Dの実体

は、その内容が検索可能でさえあれば、どこに存在していてもよい。

【0037】(2) ユーザ文書中からのキーワード抽出  
キーワード抽出部12がユーザ文書Y Dからキーワードを抽出し且つ重み付けして、プロファイル生成部16によりプロファイルに登録する（ステップS2）。このステップS2のキーワード抽出部12によるプロファイル抽出／登録処理の詳細を図3に示す。

(2-a) 単語の出現頻度によるキーワード抽出

全てのユーザ文書Y Dについて、形態素解析を行い、文書中に出現する単語（主として名詞及び動詞等）とその出現頻度とを分析する。全ユーザ文書Y D中の単語及びその頻度分布に基づき、TF/IDF法による重みが閾値を超える単語をキーワードに決定する。TF/IDF法は情報検索におけるキーワードの基本尺度を求める手法として広く知られており、文書中のターム（語）の頻度を示すTF（Term Frequency）及び全文書中で当該タームがいくかに少ない文書でしか現れないかを示すIDF（Inverse Document Frequency）を用いて、タームの重みを求める。

【0038】こうして、キーワードを決定すると同時に、キーワードとなる単語にTF/IDF法による重みを付与することで、キーワード自身に重みを持たせ、優先順位をつける（ステップS21）。

【0039】(2-b) 品詞の種類による重み付け

キーワードの重み値にさらに品詞による重みを加算する（ステップS22）。ステップS22における品詞に基づく重みのレベルは、未定義単語（形態素解析において、形態素解析用の辞書に定義されておらず品詞不明の単語）、名詞、動詞の順とする。

【0040】(2-c) 文書構造に着目した重み付け

次に、先に抽出されたキーワードに文書構造に基づいて重みを加算する（ステップS23）。例えば、SGML（Standard Generalized Markup Language：汎用マークアップ言語規約：文書整形言語の一つ）のような構造化文書等では、表題又はセクションのタグの中身より、またフォーマットされた文書からはセンタリング又は項番（例えば、“(1)、1、1、1、(a)”）等に基づいて見出し文を抽出し、それらの中に先に抽出されたキーワードが存在する場合には、当該キーワードにさらに重みを加算する。また、ワープロ（ワードプロセッサ）文書などで標準フォントと異なるフォント（例えば、強調フォント）で記述された部分に、先に抽出されたキーワードが含まれる場合には、そのキーワードにさらに重みを加算する。

【0041】このようにして得られたキーワード及び重み付けをプロファイル生成部16によりプロファイルに登録し（ステップS24）、図2のルーチンに戻る。

【0042】(3) キーワードのプリミティブ（原始的意味）への展開

こうして抽出されたキーワードについて、後のシソーラス検索の可能性を考慮し、シソーラス辞書よりブリミティブを抽出して、これを当該キーワードに属性として添付する(ステップS3)。

#### 【0043】(4) 文書種類・メディアの特定

メディア抽出部13が、ユーザ文書Y Dのメディアの分布を調べ(ステップS4)、メディアの種類が極端に偏る場合には、メディアの種類をプロファイル生成部16によりプロファイルに登録する(ステップS5)。また、文書種類抽出部14が、文書中に含まれる情報(T、X等のテキストフォーマットのスタイルファイルや、定型文書に含まれる固定文字列、MS-Word (Microsoft Word)に見られるプロパティなど)から文書の種類を特定することができるかを調べ(ステップS6)、文書の種類も特定できる場合には、文書の種類をプロファイル生成部16によりプロファイルに登録する(ステップS7)。

#### 【0044】(5) ユーザが事前に使用したカテゴリ名の登録(キーワードの分類)

カテゴリ検出部15が、ユーザがユーザ文書を整理するために事前にフォルダなどの階層構造を作成し管理していた場合の階層構造を調べ(ステップS8)、該階層構造が検出された場合には、該階層構造を採用して先に抽出されたキーワード群をカテゴリ化する(ステップS9)。その後、このようにして取得したプロファイルを格納する(ステップS10)。

【0045】なお、このプロファイル取得システムは、システムの起動時又はアイドル時に所定のタイミングで自動的に起動するようにすれば、ユーザが全く意識することなく、自動的にユーザ毎のプロファイルを取得することができる。また、ユーザ文書Y Dは、ユーザが作成したユーザ作成文書Y D1、ユーザが引用したユーザ引用文書Y D2及びユーザが参照したユーザ参照文書Y D3の全てを含んでいなくとも、これらのうちの少なくともいずれかを含んでいればよく、これら以外でもユーザの興味及び嗜好等に関連すると思われる文書であればどのような文書を含んでいてもよい。

【0046】このようにして、ユーザ作成文書Y D1、ユーザ引用文書Y D2及びユーザ参照文書Y D3のいずれかを含むユーザ文書Y Dを検索し、該ユーザ文書Y Dから出現頻度に基づいてキーワードを抽出して、該キーワードに基づいて該ユーザに対応するプロファイルを作成し、個人向けの情報提供を行うシステムに登録するためのユーザ個人の適切なプロファイルを自動的に取得する。したがって、ユーザ個人の興味及び嗜好等を示す個人情報であるプロファイルを、ユーザ自身が意識することなく、自動的に生成し取得することができ、ユーザ毎の個人向けの情報提供を可能とする情報提供システムに事前に供給することができる。

【0047】すなわち、プロファイルが自動的に生成さ

れ、取得されるので、プロファイル登録作業の負担を大幅に軽減することができる。また、ユーザ文書Y Dからプロファイルを生成するので、ユーザの興味及び嗜好に応じたプロファイルをスムーズに取得することができる。

【0048】次に、上述したプロファイル取得システムを採用したこの発明の第2の実施の形態による情報提供システムについて説明する。図4は、この発明の第2の実施の形態に係る情報提供システムの構成を模式的に示している。

【0049】図4に示す情報提供システムは、ネットワークNW等の通信系を介して互いに結合された検索端末RC及び情報提供サーバSを有して構成される。検索端末RCは、通常の場合1つの情報提供サーバSに対して複数個用意されるクライアント端末であり、上述したプロファイル取得システムにより、ユーザのプロファイルを取得して情報提供サーバSに転送するとともに、例えばユーザの操作に基づいて検索要求を発生するなどして、情報提供サーバSから提供された情報を表示するなどしてユーザに提示する。

【0050】該検索端末RCは、プロファイル取得部1、検索処理部2及び情報提示部3を有する。プロファイル取得部1は、上述したこの発明の第1の実施の形態に係るプロファイル取得システムを採用して、ユーザのプロファイルを取得する。したがって、このプロファイル取得部1は、実質的に図1と同様に構成される。なお、プロファイル取得部1は、検索端末RCの起動時又はアイドル時に所定タイミングで起動されて、ユーザ毎のプロファイルを取得する。

【0051】検索処理部2は、例えばユーザの操作に基づいて、情報検索要求を発生し、該検索要求を情報提供サーバSに送出するなど、ユーザによる情報検索に係る処理を実行する。情報提示部3は、検索処理の結果又はその他の処理によって、情報提供サーバSから提供された情報を、ディスプレイ装置に表示するなどしてユーザに提示する。

【0052】情報提供サーバSは、検索端末RCに情報を提供するサーバであり、定期的に、又は検索端末RCからの要求に従って、検索端末RCから転送されたプロファイルに対応する情報を検索して、検索端末RCを介してユーザに提供する。

【0053】該情報提供サーバSは、プロファイル格納部4、ユーザ用データベース処理部5、新規データ確認処理部6及びデータベース7を有して構成される。プロファイル格納部4は、検索端末RCのプロファイル取得部1から転送されたユーザ毎のプロファイルを格納する。ユーザ用データベース処理部5は、検索端末RCの検索処理部2から与えられた検索要求に従い、プロファイル格納部4に格納されたユーザ毎のプロファイルを用いてデータベース7を検索し、検索結果に基づく情報を

検索端末RCの検索処理部2に返す。また、ユーザ用データベース処理部5は、新規データ確認処理部6から与えられた新規データ確認要求に従い、プロフィール格納部4に格納されたユーザ毎のプロフィールを用いてデータベース7を検索し、検索結果に基づく情報を新規データ確認処理部6に返す。

【0054】新規データ確認処理部6は、予め設定された一定時間毎にユーザ用データベース処理部5に新規データ確認要求を与え、データベース7における新規データを検索させる。新規データ確認処理部6は、ユーザ用データベース処理部5により、データベース7内にユーザプロフィールに適合する新規データが検出されたときは、新規データをアクセスするためのURL等のリンク情報又は該新規データの実体を、検索端末RCの情報提示部3に、電子メール又は転送データとして送信する。データベース7は、提供すべき共有情報を格納している。

【0055】次に、図5、図6及び図7に示すフローチャートを参照して、このような情報提供システムの動作を説明する。図4には、情報検索用の検索端末RCと情報提供サーバISからなるクライアント・サーバモデルとして示している。図4では、ユーザは一つの検索端末RCで文書作成及び情報検索を行っている構成を想定している。この個人向け情報の提供を行う情報提供システムでは、次のようなサービスを実施するものと仮定する。

【0056】・プロフィールの条件に合致した文書の表示、

・プロフィールの条件に合致した新着文書の通知、  
・インタラクティブな検索（全文検索など）、及び  
・プロフィールによって必要と判断された文書メディアの変換。

【0057】検索端末RCの処理の動作を示すフローチャートを図5に示す。検索端末RCは、起動時もしくはアイドル時に所定のタイミングで予め指定したディレクトリ配下等を対象として、プロフィール取得部1によるプロフィール取得処理を実行させる（ステップS51）。

【0058】この場合、検索端末RCのOS（Operating System〜オペレーティングシステム）が、例えばウィンドウズ95（Windows 95〜Windowsは商標（Microsoft社））等である場合には、自動起動用のスタートアップフォルダ、すなわち典型的なシステムでは「C:\Windows\スタートアップ」で示されるパス、の配下にプロフィール取得部1を構成するプロフィール取得プログラムを置く。また、UNIX等のOSの場合には、クローン等にプロフィール取得プログラムが定期起動するように設定しておく。このようにして、ユーザにプロフィール取得プログラムの起動を意識させないような設定が効果的である。

【0059】次に、情報検索時に、該検索端末RCを情報提供サーバISに接続する際に、ステップS51で取得したプロフィールを情報提供サーバISに転送するように情報検索プログラム等に設定する。インターネットにおけるWWW（World WideWeb）の普及により、大半の情報共有／提供システムでは情報検索用プログラムとしてWebブラウザ（例えば、Netscape CommunicatorのNetscape Navigator（Netscape Communications社）、あるいはInternet Explorer（Microsoft社）等）を採用していることから、この発明を導入するシステムも、この形式を採用することができる。Webブラウザを用いる場合、Webブラウザの設定により、情報提供サーバISへの接続時に、ユーザ名及びパスワード等を指定して検索サービスを利用できるように、情報提供サーバISを設定する。この際に、プロフィールもクライアントRCから情報提供サーバISに転送するようにプロフィール転送用モジュールも組み込んでおく。

【0060】したがって、クライアントRCを情報提供サーバISに接続した時には（ステップS52）、クライアントRCは情報提供サーバISにプロフィールを転送する（ステップS53）。なお、後述するように、情報提供サーバISから新規文書が電子メール等により通知されている場合には（ステップS54）、情報提示部3により、ディスプレイに新規文書のリンク情報又は新規文書を表示するなどしてユーザに提示する（ステップS55）。また、ユーザにより、検索端末RCの検索処理部2を介して情報検索が要求されている場合には（ステップS56）、情報提供サーバISから返された検索結果を情報提示部3を介してユーザに提示する（ステップS57）。これらステップS54〜S57の処理は、検索端末RCが起動中繰り返し行われる。

【0061】情報提供サーバISは、クライアントRCからプロフィールを受理し、プロフィールに基づく個人向け情報提供サービスを提供する。このため、情報提供サーバISは、予め次のような準備をしておく。

【0062】まず、情報提供サーバISは、情報を蓄積するデータベース7に次のような情報を登録する。

・情報（文書）のメディア、  
・情報（文書）の種類（これはシステムに特化した基準になる）、及び  
・情報（文書）を含むテキスト。

【0063】これらを登録したデータベース7に対し、ユーザが接続した際、ユーザ環境で上述したロジックで自動取得されたプロフィールを取り込み、データベース7内の情報（文書）に重み付けを行う。プロフィールから抽出される情報は、キーワード群、メディア及び文書種類の3種類の情報であるので、文書の属性であるメディア及び種類に重み付けを行い、キーワード群を含む文書自身に重み付けを行う。

【0064】こうした準備を行った後、ユーザの検索要



求に対して前述のプロファイルによる重み付けを付加した検索を行うことにより、検索精度が向上すると考えられる。

【0065】このような、情報提供サーバISにおける所定時間毎に定期的に行われる定時処理のフローチャートを図6に示す。情報提供サーバIS側では、予め設定したタイミングで、ユーザが前回接続した際にプロファイルを受信し、プロファイル格納部4に格納したことが確認されると(ステップS61)、該プロファイルに基づいて、プロファイル情報を登録した個人用のユーザデータベースを構築する(ステップS62)。新たにプロファイルを受信されていない場合、ステップS62の処理はバイパスする。

【0066】そして、新規データ確認処理部6が、ユーザ用データベース処理部5を介して各ユーザ毎のプロファイルの条件に合致する文書を検索し(ステップS63)、プロファイルの条件を満たす文書が新規に追加されたかどうかを判定する(ステップS64)。新規文書が追加されたことが確認された場合、情報提供サーバIS側で登録済みのユーザへ電子メール等の通知・連絡系を用いて、追加文書のリンクの一覧情報又は新規文書の実体を送付する(ステップS65)。

【0067】これにより、検索端末RCを介してユーザのプロファイルに適合した情報が提供され、ユーザは興味のある情報をタイムリーに取得することができる。

【0068】また、情報提供サーバISは、ユーザ接続時にそれに応答して、ユーザの検索要求等に基づく情報の提供も行う。このユーザ接続時の情報提供サーバISの処理のフローチャートを図7に示す。

【0069】情報提供サーバISに接続したユーザが特定されると(ステップS71)、そのユーザのプロファイル情報より情報提供サーバIS内のデータベース7に登録されているインスタンス例えば文書の重みを動的に変化させる(ステップS72)。ユーザが検索端末RCの検索インタフェースである検索処理部2を介して検索要求を発行した際に(ステップS73)、ユーザ用データベース処理部5により、これら検索機能に動的に重み付けされた文書の重み値を考慮した検索を行う(ステップS74)。

【0070】例えば、全文検索の場合、検索に指定した単語すなわちタームが、文書中にどのくらいの割合で含まれるかあるいはどのくらいの頻度で出現するかに着目した重み付け検索を行い、この検索結果に対してさらにプロファイルによる重みを付加して、検索結果に優先度として反映させる。また、データベース7に登録されている文書が膨大である場合、検索時間を考慮し、プロファイルによって重み値を与えられた文書のみを対象として、検索対象を絞り込んだ絞り込み検索を実行することにより、検索対象空間を限定して検索時間を短縮する。

【0071】検索結果は、情報提供サーバISのユーザ

用データベース処理部5から検索端末RCの検索処理部2及び情報提示部3を介してユーザに通知される(ステップS75)。

【0072】すなわち、検索端末RCでは、文書の作成及び文書等の情報の検索に加えて、プロファイル情報の自動取得が行われ、情報提供サーバISに対して、ユーザのプロファイルを転送し、情報の検索を要求する。一方、情報提供サーバISでは、プロファイルによる文書データベース内の文書重み値の動的変更、プロファイルデータベース管理、文書データベース管理、重み付け全文検索及びプロファイルを加味した検索等が行われる。その結果、情報提供サーバISは、検索端末RCを介してユーザに、プロファイルで指定された条件に合う文書を提示し、プロファイルで指定された条件に合う新着情報の登録を通知し、且つプロファイルを加味した検索結果を提示する。

【0073】このようにして、ユーザ文書に基づいてユーザに対応するプロファイルを取得し、該プロファイルに基づき、又は該プロファイルに基づく重み付けを行って、情報を検索して前記ユーザに提供する。すなわち、プロファイルを自動的に取得することにより、プロファイル登録作業の負担を大幅に軽減することができ、ユーザ文書を用いるプロファイル取得により、ユーザの興味及び嗜好を的確に把握することができる。その結果として、情報検索における検索精度、つまり適合率を向上することができ、ユーザの興味に沿う文書情報の更新をタイムリーに通知することができる。

【0074】なお、この発明は上記実施の形態に限定されず、種々の変形及び応用が可能である。例えば、第1の実施の形態のプロファイル取得システムで取得したプロファイルを手動で、又は、メールで情報提供システムに送信する等してもよい。

【0075】なお、この発明のプロファイル取得システム及び情報提供システムは、専用のシステムとして構成することなく、通常のコンピュータシステムを用いて実現することができる。例えば、コンピュータシステムに上述の動作を実行するためのプログラムを格納した媒体(フロッピーディスク、CD-ROM等)から該プログラムをインストールすることにより、上述の処理を実行するシステムを構築することができる。インストールによって、当該プログラムは、コンピュータシステム内のハードディスク等の媒体に格納されて、システムを構成し、実行に供される。

【0076】また、コンピュータにプログラムを供給するための媒体は、狭義の記憶媒体に限らず、通信回線、通信ネットワーク及び通信システムのように、一時的且つ流動的にプログラム等の情報を保持する通信媒体等を含む広義の記憶媒体であってもよい。

【0077】例えば、インターネット等の通信ネットワーク上に設けたFTP(File Transfer Protocol)サー

に当該プログラムを登録し、FTPクライアントにネットワークを介して配信してもよく、通信ネットワークの電子掲示板(BBS: Bulletin Board System)等に該プログラムを登録し、これをネットワークを介して配信してもよい。そして、このプログラムを起動し、OS(Operating System)の制御下において実行することにより、上述の処理を達成することができる。さらに、通信ネットワークを介してプログラムを転送しながら起動実行することによっても、上述の処理を達成することができる。

【0078】

【発明の効果】以上説明したように、この発明によれば、個人向けの情報提供を行うシステムに登録するためのユーザ個人の適切なプロフィールを容易に取得し、該プロフィールの登録によりユーザ個人の興味及び嗜好等に適切に関連する情報の提供を可能とするプロフィール取得システム、情報提供システム、プロフィール取得方法及び媒体を提供することができる。

【図面の簡単な説明】

【図1】この発明の第1の実施の形態に係るプロフィール取得システムの構成を模式的に示すブロック図である。

【図2】図1のプロフィール取得システムにおける動作を説明するためのフローチャートである。

【図3】図2におけるキーワード抽出/登録処理を詳細に説明するためのフローチャートである。

【図4】この発明の第2の実施の形態に係る情報提供シ

\* システムの構成を模式的に示すブロック図である。

【図5】図1の情報提供システムにおける検索端末の動作を説明するためのフローチャートである。

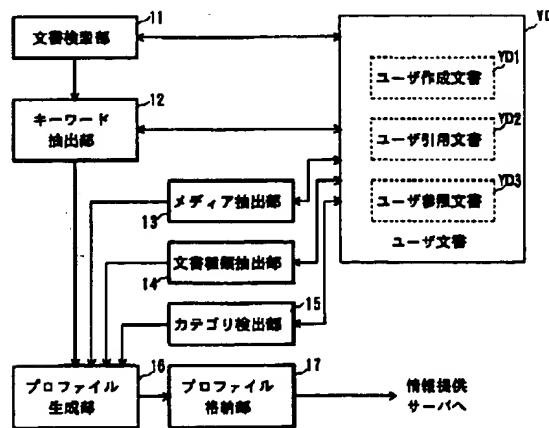
【図6】図1の情報提供システムにおける情報提供サーバの定時処理の動作を説明するためのフローチャートである。

【図7】図1の情報提供システムにおける情報提供サーバのユーザ接続時の処理動作を説明するためのフローチャートである。

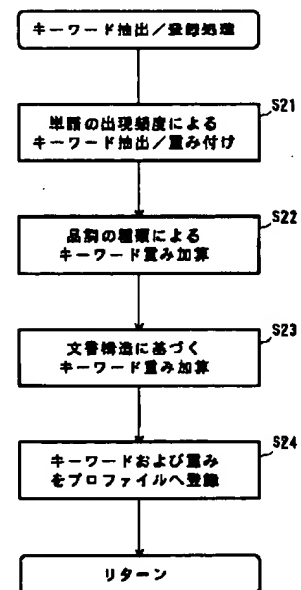
10 【符号の説明】

RC	検索端末(クライアント)
IS	情報提供サーバ
NW	ネットワーク
1	プロフィール取得部
2	検索処理部
3	情報提示部
4, 17	プロフィール格納部
5	ユーザ用データベース処理部
6	新規データ確認処理部
7	データベース
11	文書検索部
12	キーワード抽出部
13	メディア抽出部
14	文書種類抽出部
15	カテゴリ検出部
16	プロフィール生成部

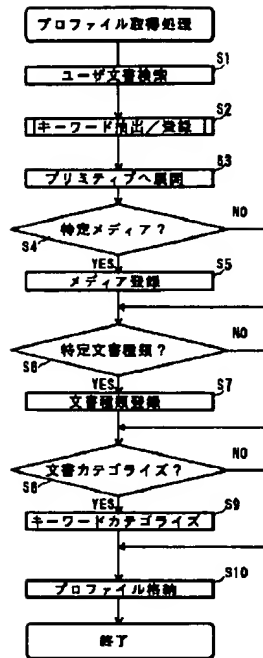
【図1】



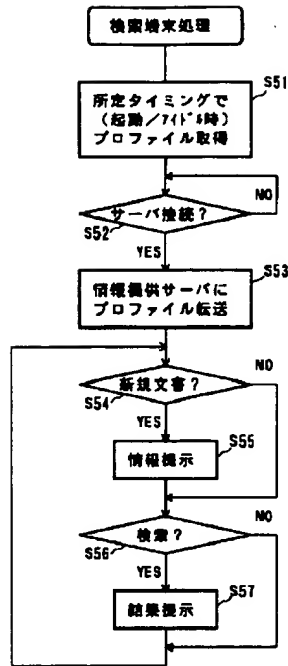
【図3】



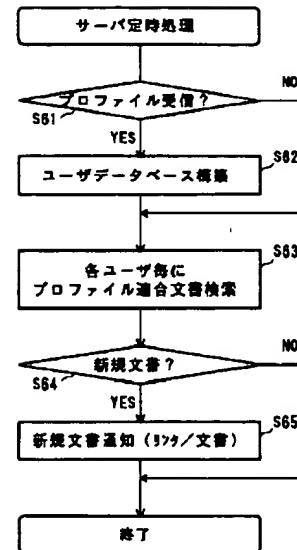
【図2】



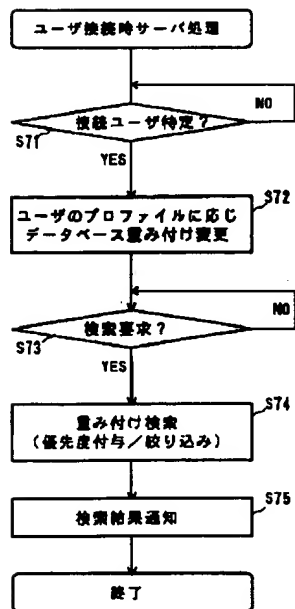
【図5】



【図6】



【図7】



〔図4〕

